American Journal of Epidemiology
Vol. 159, No. 3

Printed in U.S.A.
DOI: 10.1093/aje/kwh036

Fraction of Cases of Acquired Immunodeficiency Syndrome Prevented by the Interactions of Identified Restriction Gene Variants

M. J. Silverberg[1], M. W. Smith[2], J. S. Chmiel[3], R. Detels[4], J. B. Margolick[5], C. R. Rinaldo[6], S. J. O'Brien[7], and A. Muñoz[1]

[1] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.
[2] Basic Research Program, Science Applications International Corporation, Frederick, MD.
[3] Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL.
[4] Department of Epidemiology, School of Public Health, University of California, Los Angeles, Los Angeles, CA.
[5] Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.
[6] Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.
[7] Laboratory of Genomic Diversity, National Cancer Institute–Frederick, Frederick, MD.

Received for publication January 22, 2003; accepted for publication July 31, 2003.

Previous research has demonstrated isolated effects of host genetic factors on the progression of human immunodeficiency virus type 1 (HIV-1) infection. In this paper, the authors present a novel use of multivariable methods for estimating the prevented fraction of acquired immunodeficiency syndrome (AIDS) cases attributable to six restriction genes after accounting for their epidemiologic interactions. The methods presented will never yield a prevented fraction above 1. The study population consisted of a well-characterized cohort of 525 US men with HIV-1 seroconversion documented during follow-up (1984–1996). On the basis of a regression tree approach using a Cox proportional hazards model for times to clinical AIDS, the combinations of genes associated with the greatest protection, relative to the lack of a protective genotype, consisted of: 1) C-C chemokine receptor 5 (*CCR5*)-*Δ32* and C-C chemokine receptor 2 (*CCR2*)-*64I* (relative hazard = 0.44); 2) interleukin 10 (*IL10*)-+/+ in combination with *CCR5-Δ32* or *CCR2-64I* (relative hazard = 0.45); and 3) *IL10*-+/+ in combination with stromal-derived factor (*SDF1*)-*3′A* and *CCR5* promoter *P1*/~*P1* (relative hazard = 0.37). Overall, 30% of potential AIDS cases were prevented by the observed combinations of restriction genes (95% confidence interval: 7, 47). However, the combined effect was confined to the first 4 years following HIV-1 seroconversion. Additional research is needed to identify AIDS restriction genes with stronger and long-lasting protection to better characterize the genetic epidemiology of HIV-1.

acquired immunodeficiency syndrome; chemokines; cytokines; epidemiologic methods; HIV-1; HLA antigens; receptors, chemokine

Abbreviations: AIDS, acquired immunodeficiency syndrome; CCR2, C-C chemokine receptor 2; CCR5, C-C chemokine receptor 5; CCR5P, C-C chemokine receptor 5 promoter; CI, confidence interval; HIV-1, human immunodeficiency virus type 1; HLA, human leukocyte antigen; IL10, interleukin 10; MACS, Multicenter AIDS Cohort Study; PF, prevented fraction; RH, relative hazard; SDF1, stromal-derived factor.

Prior to the introduction of highly active antiretroviral therapy in 1996, substantial variability in rates of progression to acquired immunodeficiency syndrome (AIDS) among persons infected with human immunodeficiency virus type 1 (HIV-1) was well documented (1). It was recognized that some persons progressed to AIDS rapidly and others progressed slowly (2–4). These observations, plus the well-known fact that genetic variants influence infectious

Correspondence to Dr. Michael Silverberg, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Room E-7137, Baltimore, MD 21205 (e-mail: msilverb@jhsph.edu).

disease outcomes in animals (5–7), prompted a search for AIDS restriction genes that predicted both the likelihood of HIV-1 infection and the progression of disease once a person was infected.

Initial studies examining the influence of AIDS restriction genes on the progression of AIDS focused on the isolated effect of polymorphisms for the primary coreceptors, chemokines, cytokines, and human leukocyte antigen (HLA). A mutation of the C-C chemokine receptor 5 (*CCR5*) allele (*CCR5-Δ32*) was identified that essentially prevented infection among persons who were homozygous for the mutation (8–11) and delayed progression among those who were heterozygous for it (8, 11). Other effects found included: delayed progression among persons carrying a mutation in the C-C chemokine receptor 2 (*CCR2*) allele (*CCR2-64I*) (12); rapid progression among persons who were homozygous for the *P1* haplotype of the *CCR5* promoter (*CCR5P*) allele (*CCR5P1/P1*) (13, 14); delayed progression among persons with a mutation for stromal-derived factor (SDF1) (*SDF1 3′A/3′A*) (15); rapid progression among persons who had homozygous alleles at one, two, or three *HLA* class I loci (16); and rapid progression among persons with a polymorphism for interleukin 10 (IL10) (*IL10 +/5′A* or *5′A/5′A*) (17).

However, the prognosis of HIV-1-infected persons is likely to involve interactions of several host genes, virus genes, and other nongenetic influences. In general, genetic studies have examined the effects of AIDS restriction genes separately, although a few studies (13, 15, 17, 18) have considered the interaction of two or three genes at a time. The goal of this study was to examine the overall influence of described AIDS restriction genes on progression to AIDS among participants in the Multicenter AIDS Cohort Study (MACS) who had HIV-1 seroconversion documented during follow-up. Our approach to this question was based on regression trees, which are directly suitable for incorporating interactions among many variables. To obtain a single overall measure of the influence of genetic factors, we estimated the prevented fraction of AIDS cases, defined as the proportion of potential AIDS cases prevented by AIDS restriction genes relative to a population without a protective genotype.

MATERIALS AND METHODS

Study population, outcomes, and exposures

The MACS is an ongoing prospective cohort study of the natural history of HIV-1 infection among homosexual and bisexual men who are followed up every 6 months. A detailed description of the study has been published previously (19). Briefly, from 1984 to 1985 and from 1987 to 1991, a total of 2,195 HIV-1-seropositive and 3,427 HIV-1-seronegative men were enrolled in the MACS in four US metropolitan areas: Baltimore, Maryland; Pittsburgh, Pennsylvania; Chicago, Illinois; and Los Angeles, California. During the physical examination performed at each 6-month follow-up visit, blood samples are taken and stored at both the local sites and a national repository. We used these samples for routine HIV-1 testing by means of both enzyme-linked immunosorbent assay and Western blot assay and for assessment of the AIDS restriction genes. Since our primary aim was to assess the effect of AIDS restriction genes on HIV-1 disease progression in the natural history setting (i.e., in the absence of effective treatments prior to 1996), the study population for this analysis consisted of 525 initially HIV-1-seronegative men who seroconverted between June 1984 and July 1995 and who were followed until the diagnosis of AIDS, death, loss to follow-up, or December 31, 1995, the end of study follow-up.

The primary outcome of interest was time from HIV-1 seroconversion to the development of an AIDS-defining illness, based on category C clinical conditions listed in the Centers for Disease Control and Prevention's 1993 case definition (i.e., the immunologic criterion of a CD4-positive cell count less than 200 cells per μl was not included in our case definition) (20). Continuous surveillance data were available for each MACS participant with respect to the development of clinically defined AIDS, death, or loss to follow-up.

The exposures of interest were AIDS restriction genes known to influence the development of AIDS, based on the seminal genetic studies published prior to calendar year 2002. For each of the 525 seroconverters, we determined genetic status regarding *CCR5* (wild-type +/+ and *+/Δ32*, since *Δ32/Δ32* protects against HIV-1 infection), *CCR2* (+/+, *+/64I*, and *64I/64I*), *CCR5P* (*P1/P1*, *P1/~P1*, and *~P1/~P1*, where *~P1* represents *P2*, *P3*, or *P4*), *SDF1* (+/+, *+/3′A*, and *3′A/3′A*), *IL10* (*5′A/5′A*, *+/5′A*, and +/+), and *HLA* homozygosity (number among the *HLA-A*, *B*, and *C* loci with identical (i.e., homozygous) alleles: 2 or 3, 1, or 0). Human genomic DNA was extracted from Epstein-Barr virus-immortalized B cell lines, and stored blood specimens and specific segments were amplified by polymerase chain reaction for the determination of genotypes as previously described (8, 12, 13, 15–17). For *HLA* (16), a panel of primers specific for the *HLA-A*, *B*, and *C* loci was used to identify homozygous alleles. Even though, in principle, the *HLA A*, *B*, and *C* loci are separate loci, we used here the combination put forward by Carrington et al. (16), and hereafter we refer to six AIDS restriction genes: *CCR5*, *CCR2*, *CCR5P*, *SDF1*, *IL10*, and *HLA*.

The study protocols were reviewed and approved by the institutional review boards of the study sites. All participants provided written informed consent.

Statistical methods

The outcome for this analysis was time to AIDS from HIV-1 seroconversion. Since participants in the MACS are followed at 6-month intervals, exact dates of seroconversion were unknown. Furthermore, there were persons who missed semiannual visits between their last negative and first positive tests. To be in consonance with the demonstrated downward trend of HIV-1 incidence over time (21), we assigned a seroconversion date for these persons at one third of the time between the last HIV-1-seronegative study visit and the first HIV-1-seropositive study visit (22). To appropriately incorporate in the analysis those subjects with long times to the first positive visit, we used the time from the assigned seroconversion date to the first HIV-1-seropositive visit as the

time of entry into the observed risk sets (i.e., staggered/late entries) (23). Censored observations (i.e., AIDS-free at the last time seen) resulted from losses to follow-up, deaths unrelated to HIV-1, and freedom from AIDS at the date of analysis, which was preset as December 31, 1995 (i.e., before the introduction of highly active antiretroviral therapy).

To quantify the protective effect of each of the AIDS restriction genes, we first computed univariate relative hazards using Cox proportional hazards models (24). Persons with the genotype associated with the highest susceptibility to AIDS on the basis of previous reports (8, 11–17) were selected as the reference group. The multivariable analysis consisted of the construction of a regression tree in two stages to incorporate the known relations of the AIDS restriction genes.

The first stage of the regression tree incorporated only *CCR2*, *CCR5P*, and *CCR5*, which are structurally related and constitute the *CCR2.CCR5P.CCR5* superlocus (13, 25). These were the first AIDS restriction genes to be described, and their joint effects have been replicated in the literature (13, 14). Therefore, the first split of the regression tree was defined by the *CCR2.CCR5P.CCR5* superlocus, which has four haplotypes (i.e., [+.*P1*.+], [+.*P1.Δ32*], [*64I.P1*.+], and [+.~*P1*.+]) corresponding to nine potential genotypes for HIV-1-infected persons (those with *CCR5-Δ32/Δ32* are protected against HIV-1 infection). A full Cox regression model yielded the relative hazards for all genotypes relative to those associated with the most susceptibility to AIDS: [+.*P1*.+]/[+.*P1*.+]. To combine genotypes associated with similar risks of AIDS, we used a recursive amalgamation algorithm (26, 27) whereby two categories were joined if they yielded the lowest deviance (i.e., goodness-of-fit likelihood ratio statistic relative to the model with the two categories separated) among persons with deviances below 1.32, which corresponds to the 75th percentile of the chi-squared test with 1 df.

For each resulting node of the superlocus, the second stage of the regression tree analysis consisted of identifying subsequent branches of the tree defined by the three remaining AIDS restriction genes (i.e., *SDF1*, *IL10*, and *HLA* homozygosity) using standard binary recursive partitioning methodology (26–30). Specifically, for the three genotypes (x_1, x_2, x_3) of each AIDS restriction gene (e.g., for *SDF1*, x_1 = *3′A/3′A*, x_2 = +/*3′A*, and x_3 = +/+), we fitted two Cox regression models to identify the nature of the association as dominant (i.e., including as a covariate an indicator for x_1 or x_2) or recessive (i.e., including as a covariate an indicator for x_1). We used the likelihood ratio statistic as the dissimilarity measure, and a node was split if the resulting nodes contained more than 10 persons and if the largest likelihood ratio statistic was above 2.07, which corresponds to the 85th percentile of a chi-squared test with 1 df.

Subsequent splits for newly defined nodes were determined in the same way, including the possible determination of an association as codominant (i.e., including as a covariate an indicator for x_1 for nodes defined by x_1 or x_2, or including as a covariate an indicator for x_2 for nodes defined by x_2 or x_3). Once the full tree was derived, we fitted a Cox regression model to the full data to obtain relative hazards using the described reference group. We identified final nodes by combining nonreference nodes with similar relative hazards using the amalgamation procedures described. Kaplan-Meier survival curves for the final nodes and relative hazards (RH_i) were computed with the reference group coded by $i = 0$ (i.e., $RH_0 = 1$).

A multivariate relative hazard (RH_M) was computed as a weighted average (based on weights determined by the percentage of seroconverters (p_i) at each final node) of the RH_i's for all nonreference final nodes (i.e., $RH_M = \Sigma p_i\, RH_i / \Sigma p_i$ with summations for $i > 0$). On the basis of both the RH_M and the total percentage with protective genotypes (i.e., Σp_i with summations for $i > 0$), we computed the multivariate prevented fraction (PF_M) as

$$\Sigma p_i \times (1 - RH_M) = (1 - p_0) \times (1 - RH_M).$$

The PF_M is interpreted as the proportion of all potential AIDS cases that were prevented as a result of the studied AIDS restriction genes. An added feature of our regression tree approach (in contrast to ordinary multiple regression) is that the joint effects of the AIDS restriction genes will never yield a prevented fraction above 1. In addition, if the AIDS restriction genes are indeed protective (i.e., $RH_M < 1$), the prevented fraction will be above zero; otherwise, it can take negative values. Negative values of the prevented fraction will be indicative of no protection.

A confidence interval for the PF_M was obtained using the delta method for the log of RH_M as a function of β_i for $i > 0$. Specifically,

$$\text{Var}[\log RH_M] = \Sigma\Sigma\, d_i\, d_j\, c_{ij},$$

where $d_i = p_i \exp(\beta_i)/((1 - p_0)RH_M)$ and c_{ij} is the covariance between β_i and β_j obtained from a Cox regression on the final nodes. A 95 percent confidence interval for PF_M is given by

$$(1 - p_0) \times (1 - RH_M \exp(\pm 1.96[\text{Var}(\log RH_M)]^{1/2})).$$

To assess the robustness of the delta method when applied to our data, we repeated the second stage of the analysis on 100 bootstrap samples (i.e., random sample with replacement from the 525 seroconverters), allowing the tree to vary with each bootstrap for all nonreference nodes of the superlocus. We compared the third and 98th of the 100 ordered PF_M's with the 95 percent confidence interval obtained using the delta method in the original sample of the 525 seroconverters.

To determine the prevented fraction at different times since seroconversion and to allow for departures from the proportional hazards assumption over the full time span, we performed the final analysis in strata of years of follow-up defined by 0.0–4.0, 4.1–8.0, and 8.1–12.0 years, with the use of staggered entries (23) for the last two strata. The 95 percent confidence interval for each stratum was computed using the delta method as described above.

RESULTS

By the end of study follow-up, 218 (41.5 percent) of the 525 HIV-1 seroconverters had been diagnosed with AIDS

TABLE 1. Descriptive data for 525 human immunodeficiency virus type 1 seroconverters from the Multicenter AIDS* Cohort Study, 1984–1996

Characteristic	Median	IQR*	No. (%)
Date of seroconversion	September 1986	April 1985–September 1989	
Time (years) between last seronegative and first seropositive study visit	0.5	0.5–0.6	≤1 year: 84%; ≤4 years: 95%
Age (years) at seroconversion	33	28–39	
Duration (years) of follow-up	6.2 (maximum = 11.5)	4.1–8.6	
Diagnosed with AIDS			218 (41.5%)
Died			174 (33.1%)

* AIDS, acquired immunodeficiency syndrome; IQR, interquartile range.

(table 1) with 49, 128, and 41 diagnoses occurring at ≤4 years, 4.1–8.0 years, and 8.1–12.0 years, respectively. Of the 307 persons not observed to develop AIDS, censored observations included 20 (6.5 percent) persons who died during follow-up without an AIDS diagnosis, 16 (5.2 percent) who were lost to follow-up, and 271 (88.3 percent) who were AIDS-free as of December 31, 1995. A complete genetic profile on the six AIDS restriction genes was available for 96 percent of all seroconverters. Twenty-three seroconverters had missing data on *HLA* class 1 loci but had complete data on the other five AIDS restriction genes.

In agreement with previous studies, which prominently included the data on MACS seroconverters, genotypes that were expected to delay the onset of AIDS were associated with protective relative hazards in the univariate analyses (table 2). Homozygosity at *SDF1* (*3′A/3′A*) was the only genotype with a statistically significantly ($p < 0.05$) reduced hazard of AIDS (RH = 0.24; $p = 0.013$). Previous investigators found significant p values in some individual cohorts, usually conclusively only after combining several studies (25, 31).

Table 3 shows the results of joint analysis of the *CCR2.CCR5P.CCR5* superlocus. Since only seven persons had the *CCR2-64I/64I* genotype, they were combined with persons who had the *CCR2-+/64I* genotype, giving us eight categories in table 3 containing HIV-1 seroconverters. The presence of categories with no seroconverters was attributable to linkage disequilibrium. Out of all of the categories, persons with both the *CCR5-Δ32* polymorphism and the *CCR2-64I* polymorphism were the most protected (RH = 0.44), which is consistent with an independent and additive effect of these two alleles as previously reported (12). Using the amalgamation procedure described, we created a tree with three nodes, as indicated by the solid boxes in table 3. To separate the distinct effects of *~P1/~P1* for the promoter and the joint effect of *CCR5-Δ32* and *CCR2-64I*, we split the largest node of the amalgamated tree into three nodes, as indicated by the dashed boxes in table 3.

The five final categories of the superlocus from the first stage of the analysis formed the first split of the regression tree (see split A in figure 1). The second stage of the analysis used standard binary recursive partitioning methodology with a likelihood ratio statistic greater than 2.07 as the splitting criterion. The resulting tree, with eight terminal nodes depicted in squares along with the Kaplan-Meier curves of the groups at each split, is shown in figure 1. The Kaplan-Meier curves corresponding to the split defined by *IL10*

TABLE 2. Relative hazard of progression to AIDS* among 525 human immunodeficiency virus type 1 seroconverters from the Multicenter AIDS Cohort Study in univariate analyses, according to the possession of six AIDS restriction genes, 1984–1996

Genotype	No.	%	% with AIDS	RH*	*p* value
*CCR5**					
+/+	438	83.4	41.3	1	
+/Δ32	87	16.6	42.5	0.76	0.119
*CCR2**					
+/+	421	80.2	42.8	1	
+/64I	97	18.5	36.1	}	
64I/64I	7	1.3	42.9	} 0.77	0.135
*CCR5P**					
P1/P1	152	29.0	46.7	1	
P1/~P1	265	50.5	41.5	0.94	0.698
~P1/~P1	108	20.5	34.3	0.83	0.349
*SDF1**					
+/+	344	65.5	41.6	1	
+/3′A	163	31.0	44.2	1.05	0.715
3′A/3′A	18	3.4	16.7	0.24	0.013
*IL10**					
5′A/5′A	37	7.0	54.0	1	
+/5′A	210	40.0	44.8	0.89	0.623
+/+	278	53.0	37.4	0.70	0.148
*HLA** homozygosity					
2 or 3	27	5.4	44.4	1	
1	103	20.5	42.7	0.63	0.153
0	372	74.1	42.2	0.59	0.079
Missing data	23		21.7		

* AIDS, acquired immunodeficiency syndrome; RH, relative hazard; CI, confidence interval; CCR5, C-C chemokine receptor 5; CCR2, C-C chemokine receptor 2; CCR5P, C-C chemokine receptor 5 promoter; SDF1, stromal-derived factor; IL10, interleukin 10; HLA, human leukocyte antigen.

TABLE 3. Relative hazard of progression to AIDS* among 525 human immunodeficiency virus type 1 seroconverters from the Multicenter AIDS Cohort Study, according to possession of the *CCR5*,* *CCR2*,* and *CCR5P genotypes, 1984–1996†**

CCR5 genotype	*CCR2* genotype	*CCR2.CCR5P.CCR5* superlocus‡			*CCR5* and *CCR2* genotypes only
		CCR5P genotype			
		P1/P1	*P1/~P1*	*~P1/~P1*	
+/+	+/+	1§ (*n* = 75)	0.76 (*n* = 163)	0.60 (*n* = 108)	1§ (*n* = 346)
+/*Δ32*	+/+	0.57 (*n* = 24)	0.56 (*n* = 51)	(*n* = 0)	0.74 (*n* = 75)
+/+	+/*64I* or *64I*/*64I*	0.56 (*n* = 41)	0.58 (*n* = 51)¶	(*n* = 0)	0.75 (*n* = 92)
+/*Δ32*	+/*64I*	0.44 (*n* = 12)	(*n* = 0)	(*n* = 0)	0.59 (*n* = 12)

* AIDS, acquired immunodeficiency syndrome; CCR5, C-C chemokine receptor 5; CCR2, C-C chemokine receptor 2; CCR5P, C-C chemokine receptor 5 promoter.

† The solid box indicates the results of amalgamation of cells if the likelihood ratio statistic is less than 1.32 (75th percentile of the chi-squared test with 1 df). The dashed boxes indicate categories used for the analysis.

‡ Categories correspond to genotypes defined by four haplotypes for the *CCR2.CCR5P.CCR5* superlocus: [+.*P1*.+], [+.*P1*.*Δ32*], [*64I*.*P1*.+], and [+.~*P1*.+] (13, 25).

§ Reference category.

¶ Includes seroconverters with the *CCR2*-+/*64I* genotype but not the *CCR2*-*64I*/*64I* genotype.

diverged over time, supporting later effects (17). The 75 seroconverters with wild-type *CCR5*, wild-type *CCR2*, and the *CCR5P1P1* haplotype had the highest hazard rate for AIDS and comprised the reference group for all subsequent analyses.

Creation of the regression tree presented in figure 1 resulted in the identification of eight nodes that were not split further. Although the relative hazards ranged from 0.37 to 0.93, there were some combinations of restriction genes that were associated with similar risks of AIDS. For example, the nodes with the most protective relative hazards consisted of persons with the following combinations of genotypes: 1) *CCR5-Δ32* and *CCR2-64I* (RH = 0.44); 2) *IL10*-+/+ in combination with *CCR5–Δ32* or *CCR2-64I* (RH = 0.45); and 3) *IL10*-+/+ in combination with *SDF1*-+/*3′A* or *SDF1-3′A*/*3′A* and *CCR5P1*/*~P1* (RH = 0.37). Nodes with similar hazards were amalgamated, as indicated by dashed lines in figure 1, resulting in four final nodes, which are depicted with hexagons (the reference group and three groups with one or more protective genotypes). Based on the prevalence and relative hazard associated with the four final nodes, the overall prevented fraction was estimated to be 0.296 (95 percent confidence interval (CI): 0.072, 0.467) (table 4). The prevented fraction corresponding to the eight nodes of the regression tree in figure 1 before amalgamation was 0.298 (95 percent CI: 0.074, 0.467), indicating that the amalgamation procedure had no effect on the summary measure.

Consonant with the lack of proportionality of hazards exhibited by the curves for the final nodes in figure 2 (e.g., crossing of Kaplan-Meier curves of nodes 0 and 1), the overall effect of protective genotypes was not constant over time. We found a strong protective effect within the first 4 years of HIV-1 infection, which diminished considerably over the course of infection. The prevented fraction decreased from 0.514 in the first 4 years of infection to 0.144 between 8.1 years and 12.0 years (table 5). The 95 percent confidence intervals for the prevented fractions at 4.1–8.0 years and 8.1–12.0 years included zero. Node 3, which was the only category with a consistent protective effect in the later stages of HIV-1 infection, was also the only category that included protective genotypes for *IL10*.

DISCUSSION

In this multivariate analysis carried out among HIV-1 seroconverters participating in the MACS, we determined that six loci, each of which has a modest effect on the development of AIDS when examined in isolation, prevented 30 percent (95 percent CI: 7, 47) of all potential AIDS cases when examined in combination, relative to the lack of a protective genotype. The lower 95 percent confidence limit of 7 percent is consistent with a protective effect for the combined influence of the studied AIDS restriction genes. However, the combined effect of the studied AIDS restriction genes was confined to the first 4 years following HIV-1

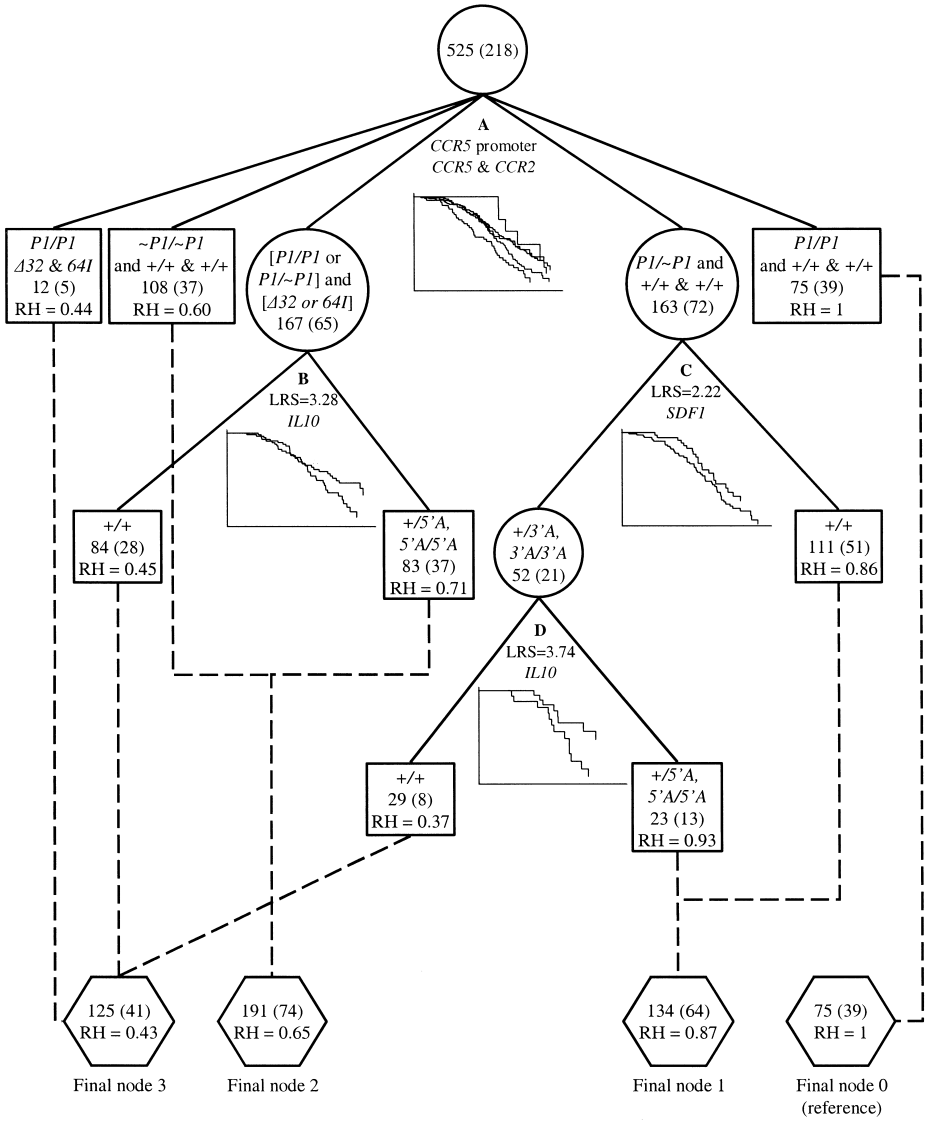

FIGURE 1. Regression tree for progression to clinical acquired immunodeficiency syndrome (AIDS) among 525 human immunodeficiency virus type 1 seroconverters from the Multicenter AIDS Cohort Study, 1984–1996. Nodes with a subsequent split defined by additional AIDS restriction genes are depicted in circles, and nodes without a subsequent split are depicted in boxes. Splits in the tree are designated A, B, C, and D. Final nodes resulting from amalgamation (dashed lines) are depicted in hexagons. Information presented for each node includes the genotypes of AIDS restriction genes that define the node, the number of seroconverters, and (in parentheses) the number of AIDS cases. CCR5, C-C chemokine receptor 5; CCR2, C-C chemokine receptor 2; RH, relative hazard; LRS, likelihood ratio statistic (for the AIDS restriction gene(s) defining a given split); IL10, interleukin 10; SDF1, stromal-derived factor.

seroconversion, while an individual effect of *IL10* was evident later.

Previous studies have addressed the issue of the interaction of genetic polymorphisms. Martin et al. (13) reported a 32 percent decreased hazard among persons with protective genotypes for *CCR5* or *CCR2* in combination with *SDF1* and a 52 percent elevated hazard among persons who were homozygous for the *CCR5P1* haplotype, in comparison with the reference group of all other combinations of protective genotypes. Winkler et al. (15) observed prolonged survival among persons with protective genotypes for *SDF1* and *CCR5* or *CCR2* in comparison with those with protective genotypes for *SDF1* alone. Finally, Shin et al. (17) identified a 38 percent reduced hazard of AIDS among persons with protective genotypes for *CCR5* or *CCR2* in combination with *IL10* as compared with those with protective genotypes for *CCR5* or *CCR2* only. In a meta-analysis, Ioannidis et al. (18) reported relative hazards of 0.74, 0.76, and 0.66 for persons with protective genotypes for *CCR5*, *CCR2*, and both, respectively, compared with persons with wild-type *CCR5*

TABLE 4. Prevented fraction of AIDS* cases conferred by combinations of five AIDS restriction genotypes among 525 human immunodeficiency virus type 1 seroconverters from the Multicenter AIDS Cohort Study, 1984–1996†

Final node‡	No.	% with AIDS	Genotype					% of sero-converters	RH*	*p* value§
			*CCR5**	*CCR2**	*CCR5P**	*SDF1**	*IL10**			
0	75	52	+/+	+/+	*P1/P1*			14.3	1	
1	134	48	+/+	+/+	***P1/~P1***	+/+		25.5	0.87	0.511
						+/3′A or ***3′A/3′A***	+/*5′A* or *5′A/5′A*			
2	191	39	***Δ32*** or ***64I***		***P1/~P1*** or *P1/P1*		+/*5′A* or *5′A/5′A*	36.4	0.65	0.028
			+/+	+/+	***~P1/~P1***					
3	125	33	+/+	+/+	***P1/~P1***	***+/3′A*** or ***3′A/3′A***	***+/+***	23.8	0.43	<0.001
			Δ32 or ***64I***		***P1/~P1*** or *P1/P1*		***+/+***			
			Δ32 and ***64I***		*P1/P1*					

Overall prevented fraction = 0.296 (95% confidence interval: 0.072, 0.467)

* AIDS, acquired immunodeficiency syndrome; CCR5, C-C chemokine receptor 5; CCR2, C-C chemokine receptor 2; CCR5P, C-C chemokine receptor 5 promoter; SDF1, stromal-derived factor; IL10, interleukin 10; RH, relative hazard.

† Boldface type in the table indicates the genotype of the AIDS restriction gene contributing protection in that node.

‡ Final node from the regression tree shown in figure 1.

§ *p* value from Cox regression analysis with indicator variables for the nodes conferring protection.

and *CCR2* alleles. The corresponding relative hazards in our data were 0.74, 0.75, and 0.59, confirming the conclusions from the meta-analysis.

Prior studies have also indicated that the protective genotypes affect different stages of HIV-1 infection. Martin et al. (13) reported an earlier effect (0–5 years) for *CCR5P*, and

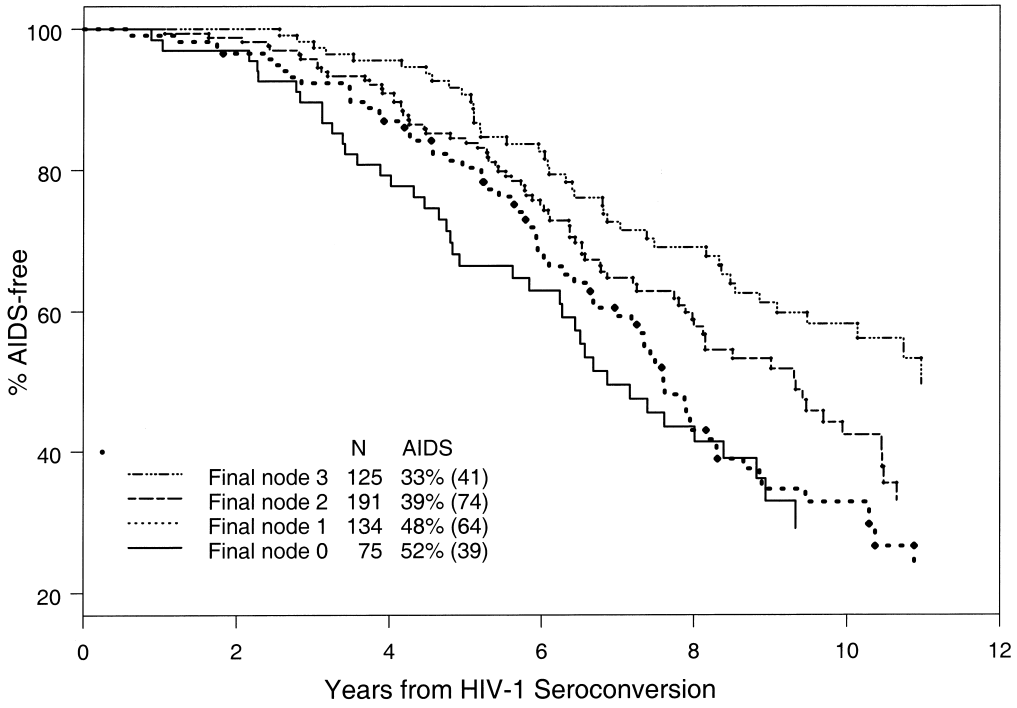

FIGURE 2. Kaplan-Meier curves (final nodes from the regression tree shown in figure 1 and table 4) for progression to clinical acquired immunodeficiency syndrome (AIDS) among 525 human immunodeficiency virus type 1 (HIV-1) seroconverters from the Multicenter AIDS Cohort Study, 1984–1996. *N*, number of seroconverters.

TABLE 5. Change in the prevented fraction of AIDS* cases conferred by AIDS restriction genes over time among 525 human immunodeficiency virus type 1 seroconverters from the Multicenter AIDS Cohort Study, 1984–1996

Final node†	Time interval (years) after seroconversion								
	≤4.0			4.1–8.0			8.1–12.0		
	No.	p‡ (%)	RH*	No.§	p‡ (%)	RH	No.¶	p‡ (%)	RH
0	75	14.3	1	60	12.7	1	33	11.3	1
1	134	25.5	0.59	117	24.9	1.03	58	20.1	1.04
2	191	36.4	0.40	175	37.0	0.74	111	38.6	0.94
3	125	23.8	0.19	120	25.4	0.54	87	30.0	0.57
Overall	525	100.0	0.40	472	100.0	0.77	288	100.0	0.84
PF_M*	0.514 (95% CI*: 0.247, 0.686)			0.204 (95% CI: –0.189, 0.467)			0.144 (95% CI: –0.934, 0.621)		

* AIDS, acquired immunodeficiency syndrome; RH, relative hazard; PF_M, multivariate prevented fraction; CI, confidence interval.

† Final node from the regression tree shown in figure 1.

‡ Percentage of seroconverters in node.

§ No. of seroconverters at baseline times the value of the node-specific Kaplan-Meier estimate at 4 years in figure 2.

¶ No. of seroconverters at baseline times the value of the node-specific Kaplan-Meier estimate at 8 years in figure 2.

Shin et al. (17) reported a later effect (>5 years) for *IL10*. Winkler et al. (15) reported stronger associations with later endpoints (i.e., the 1987 definition of AIDS and death) for the homozygous *SDF1* mutation, which is suggestive of later effects. The overall influence of genetic factors on the development of AIDS in our study was observed early in the course of infection. Of all potential AIDS cases that would have occurred within the first 4 years of infection (i.e., among rapid progressors), 51 percent were prevented by the genotype of the study participants. The overall prevented fraction beyond 4 years was substantially reduced and was not statistically significant ($p > 0.05$). Consistent with the previous findings for *IL10* (17), the node with the strongest effect beyond 4 years was node 3, which was also the only node that included protective genotypes for *IL10*. The *SDF1-3′A/3′A* genotype was only identified in 3.5 percent of seroconverters, which explains why it did not influence the overall prevented fraction in later years.

In additional analyses, we allowed the composition of the regression tree to vary in the three time periods described above, but the conclusions were unchanged (data not shown). It is also possible that our follow-up of seroconverters was not long enough to capture AIDS cases, given the median follow-up time of 6.2 years; however, 25 percent of participants were followed for 8.6 years or longer, with a maximum of 11.5 years. A more likely explanation is the emergence of HIV-1 virus populations several years after infection that were capable of overcoming the resistance provided by the combined early effect of the AIDS restriction genes.

We used the delta method to compute the standard error of the prevented fraction. It is easily derived from Cox regression analysis on the final nodes of the regression tree and does not require intensive computational methods (e.g., bootstrap methods). Furthermore, the 95 percent confidence interval obtained from the 100 bootstrap samples (95 percent CI: 0.035, 0.476) was close to the 95 percent confidence interval of the observed tree prior to amalgamation (95 percent CI: 0.074, 0.467) obtained using the delta method, with only a slightly reduced lower bound, indicating that the variability introduced by allowing the tree to vary was not substantial. The reduced precision resulting from the use of only 100 bootstrap samples may also explain the slight discrepancy. Nevertheless, the interpretation of the prevented fraction and confidence interval obtained using the delta method is conditional on the final observed tree. It would be useful to validate the prevented fraction using a different but similarly HIV-1-infected population.

While we had complete genetic data on 96 percent of the participants, 23 persons had missing data on *HLA* class 1 loci, and we were concerned that this explained why *HLA* was not included in our final tree. Therefore, we completed the missing *HLA* data on these persons based on the observed *HLA* data in persons with the same data on the other five AIDS restriction genes. With the use of standard multiple imputation methods (32), the magnitude of the univariate relative hazards for *HLA* and the composition of the final regression tree were unchanged (data not shown). Thus, the null results for *HLA* are not likely to be explained by the missing data. We used here a composite of the *HLA* data, and it is possible that using specific alleles (e.g., *B57* and *B35*) would have resulted in a refined tree.

An additional concern was the inclusion of seroconverters with longer lag times (i.e., ≥1 year) between the last HIV-1-negative and first HIV-1-positive visits ($n = 85$). The analysis appropriately accounted for the staggered/late entries of these persons, though we were also interested in the effect of confining the analysis to those with shorter lag times and thus more well-defined dates of seroconversion. The resulting prevented fraction was 0.339 (95 percent CI: 0.120, 0.505),

suggesting that the analysis based on the full data set incorporated some random error, resulting in a prevented fraction closer to zero. Since the estimates were relatively close, we chose to present the results for the complete data set.

As the field of host genetics and AIDS evolves, evaluation of the interactions between identified AIDS restriction genes becomes increasingly complex. Our approach provides a more comprehensive analysis resulting in the estimation of an easily interpretable summary measure: the prevented fraction. We also introduced an easily implemented method of calculating a confidence interval for this measure. This measure is particularly relevant for this field, since it incorporates both the strength of the association and the prevalence of a particular combination of AIDS restriction genes. Note that the prevented fraction is only generalizable to populations with a similar prevalence of restriction genes. Nevertheless, using the relative hazards reported here, which are expected to be internally valid, one can estimate the prevented fraction for a population with a different prevalence of AIDS restriction genes.

In summary, we have presented a novel use of multivariable methods for examining the influence of genetic factors on the progression of HIV-1 infection to AIDS in a well-characterized cohort of HIV-1 seroconverters. As additional AIDS restriction genes are identified, the prevented fraction can be expected to increase. Despite the considerable proportion of cases averted as a result of AIDS restriction genes, the majority of potential cases (≥70 percent) were not affected. This highlights the need to continue searching for additional genetic modifiers of the survival of HIV-1-infected persons.

ACKNOWLEDGMENTS

The Multicenter AIDS Cohort Study is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute (grants UO1-AI-35042, 5-MO1-RR-00722 (GCRC), UO1-AI-35043, UO1-AI-37984, UO1-AI-35039, UO1-AI-35040, UO1-AI-37613, UO1-AI-35041, and NO1-CO-12400).

The authors acknowledge the assistance of Richard Skolasky in data management and preliminary analysis.

The Multicenter AIDS Cohort Study (http://www.statepi.jhsph.edu/macs/macs.html) includes the following research centers: *Baltimore, Maryland*—Bloomberg School of Public Health, Johns Hopkins University: Joseph B. Margolick (Principal Investigator), Haroutune Armenian, Barbara Crain, Adrian Dobs, Homayoon Farzadegan, Nancy Kass, Shenghan Lai, Justin McArthur, and Steffanie Strathdee; *Chicago, Illinois*—Howard Brown Health Center, Feinberg School of Medicine, Northwestern University, and Cook County Bureau of Health Services: John P. Phair (Principal Investigator), Joan S. Chmiel (Co-Principal Investigator), Sheila Badri, Bruce Cohen, Craig Conover, Maurice O'Gorman, Frank Pallela, Daina Variakojis, and Steven M. Wolinsky; *Los Angeles, California*—Schools of Public Health and Medicine, University of California, Los Angeles: Roger Detels and Beth Jamieson (Principal Investigators), Barbara R. Visscher (Co-Principal Investigator), Anthony Butch, John Fahey, Otoniel Martínez-Maza, Eric N. Miller, John Oishi, Paul Satz, Elyse Singer, Harry Vinters, Otto Yang, and Stephen Young; *Pittsburgh, Pennsylvania*—Graduate School of Public Health, University of Pittsburgh: Charles R. Rinaldo (Principal Investigator), Lawrence Kingsley (Co-Principal Investigator), James T. Becker, Phalguni Gupta, John Mellors, Sharon Riddler, and Anthony Silvestre; *Baltimore, Maryland*—Data Coordinating Center, Johns Hopkins Bloomberg School of Public Health: Alvaro Muñoz (Principal Investigator), Lisa P. Jacobson (Co-Principal Investigator), Stephen R. Cole, Haitao Chu, Janet Schollenberger, Eric Seaberg, Michael Silverberg, and Sol Su; *Bethesda, Maryland*—National Institute of Allergy and Infectious Diseases: Carolyn Williams; National Cancer Institute: Sandra Melnick, Stephen J. O'Brien, and Michael W. Smith.

The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the US government.

REFERENCES

1. Muñoz A, Sabin CA, Phillips AN. The incubation period of AIDS. AIDS 1997;11(suppl A):S69–76.
2. Rutherford GW, Lifson AR, Hessol NA, et al. Course of HIV-I infection in a cohort of homosexual and bisexual men: an 11 year follow up study. BMJ 1990;301:1183–8.
3. Phair J, Jacobson L, Detels R, et al. Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: The Multicenter AIDS Cohort Study. J Acquir Immune Defic Syndr 1992;5:490–6.
4. Buchbinder SP, Katz MH, Hessol NA, et al. Long-term HIV-1 infection without immunologic progression. AIDS 1994;8:1123–8.
5. Rowe WP, Humphrey JB, Lilly F. A major genetic locus affecting resistance to infection with murine leukemia viruses. 3. Assignment of the *Fv-1* locus to linkage group 8 of the mouse. J Exp Med 1973;137:850–3.
6. O'Brien SJ, Roelke ME, Marker L, et al. Genetic basis for species vulnerability in the cheetah. Science 1985;227:1428–34.
7. Gardner MB, Kozak CA, O'Brien SJ. The Lake Casitas wild mouse: evolving genetic resistance to retroviral disease. Trends Genet 1991;7:22–7.
8. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. Science 1996;273:1856–62.
9. Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the *CCR-5* chemokine receptor gene. Nature 1996;382:722–5.
10. Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 1996;86:367–77.
11. Zimmerman PA, Buckler-White A, Alkhatib G, et al. Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. Mol Med 1997;3:23–36.

12. Smith MW, Dean M, Carrington M, et al. Contrasting genetic influence of *CCR2* and *CCR5* variants on HIV-1 infection and disease progression. Science 1997;277:959–65.
13. Martin MP, Dean M, Smith MW, et al. Genetic acceleration of AIDS progression by a promoter variant of *CCR5*. Science 1998;282:1907–11.
14. McDermott DH, Zimmerman PA, Guignard F, et al. *CCR5* promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). Lancet 1998;352:866–70.
15. Winkler C, Modi W, Smith MW, et al. Genetic restriction of AIDS pathogenesis by an *SDF-1* chemokine gene variant. Science 1998;279:389–93.
16. Carrington M, Nelson GW, Martin MP, et al. HLA and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. Science 1999;283:1748–52.
17. Shin HD, Winkler C, Stephens JC, et al. Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of *IL10*. Proc Natl Acad Sci U S A 2000;97:14467–72.
18. Ioannidis JP, Rosenberg PS, Goedert JJ, et al. Effects of *CCR5-Δ32*, *CCR2-64I*, and *SDF-1 3′A* alleles on HIV-1 disease progression: an international meta-analysis of individual-patient data. Ann Intern Med 2001;135:782–95.
19. Kaslow RA, Ostrow DG, Detels R, et al. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol 1987;126:310–18.
20. 1993 revised classification system for HIV-1 infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Morb Mortal Wkly Rep 1992;41:1–19.
21. Kingsley LA, Zhou SY, Bacellar H, et al. Temporal trends in human immunodeficiency virus type 1 seroconversion 1984–1989: a report from the Multicenter AIDS Cohort Study (MACS). Am J Epidemiol 1991;134:331–9.
22. Detels R, Muñoz A, McFarlane G, et al. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV-1 infection duration. Multicenter AIDS Cohort Study Investigators. JAMA 1998;280:1497–503.
23. Muñoz A, Hoover DR. Use of cohort studies for evaluating AIDS therapies. In: Finkelstein DM, Schoenfeld DA, eds. AIDS clinical trials. New York, NY: John Wiley and Sons, Inc, 1995:423–46.
24. Cox DR, Oates D. Analysis of survival data. New York, NY: Chapman and Hall, Inc, 1984.
25. O'Brien SJ, Moore JP. The effect of genetic variation in chemokines and their receptors on HIV-1 transmission and progression to AIDS. Immunol Rev 2000;177:99–111.
26. Ciampi A, Lawless JF, McKinney SM, et al. Regression and recursive partition strategies in the analysis of medical survival data. J Clin Epidemiol 1988;41:737–48.
27. LeBlanc M, Crowley J. Survival trees by goodness of split. J Am Stat Assoc 1993;88:457–67.
28. Breiman L, Friedman JH, Olshen RA, et al. Classification and regression trees. Monterey, CA: Wadsworth, Inc, 1984.
29. Segal MR. Regression trees for censored data. Biometrics 1988;44:35–47.
30. Zhang H, Singer B. Recursive partitioning in the health sciences. New York, NY: Springer Verlag, 1999.
31. O'Brien SJ, Nelson GW, Winkler CA, et al. Polygenic and multifactorial disease gene association in man: lessons from AIDS. Annu Rev Genet 2000;34:563–91.
32. Little RJ, Rubin DB. Statistical analysis with missing data. 1st ed. New York, NY: John Wiley and Sons, Inc, 1987.